



HAL
open science

Global sensitivity analysis of large-scale numerical landslide models based on Gaussian-Process meta modelling

Jeremy Rohmer, Evelyne Foerster

► **To cite this version:**

Jeremy Rohmer, Evelyne Foerster. Global sensitivity analysis of large-scale numerical landslide models based on Gaussian-Process meta modelling. *Computers & Geosciences*, 2011, 37 (7), pp.917-927. 10.1016/j.cageo.2011.02.020 . hal-00578781

HAL Id: hal-00578781

<https://brgm.hal.science/hal-00578781>

Submitted on 22 Mar 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Global sensitivity analysis of large-scale numerical 2 landslide models based on Gaussian-Process meta- 3 modelling

4 Jeremy Rohmer^{1*}, Evelyne Foerster¹

5 ¹BRGM, Natural Risks and Safety of CO₂ geological Storage, 3, av. Claude Guillemin BP
6 36009, 45060 Orléans Cedex 2, FRANCE

7 *e-mail: j.rohmer@brgm.fr, e.foerster@brgm.fr*

9 Abstract

10 Large-scale landslide prediction is typically based on numerical modelling, with computer
11 codes generally involving a large number of input parameters. Addressing the influence of
12 each of them on the final result and providing a ranking procedure may be useful for risk
13 management purposes. This can be performed by a variance-based global sensitivity analysis.
14 Nevertheless, such an analysis requires a large number of computer code simulations, which
15 appears impracticable for computationally-demanding simulations, with computation times
16 ranging from several hours to several days. To overcome this difficulty, we propose a “meta-
17 model”-based strategy consisting in replacing the complex simulator by a “statistical
18 approximation” provided by a Gaussian-Process (GP) model. This allows computation of
19 sensitivity measures from a limited number of simulations. For illustrative purposes, the
20 proposed methodology is used to rank in terms of importance the properties of the

*Corresponding author: J. Rohmer, e-mail: j.rohmer@brgm.fr, Tel. + 33 2 38 64 30 92, Fax. +33 2 38 64 36 89

21 elastoplastic model describing the complex behaviour of the slip surface in the La Frasse
22 landslide (Switzerland).
23 One limitation of the GP-based methodology is that the computation of sensitivity measures is
24 associated with uncertainty as the simulator is approximated using a training sample of small
25 size, *i.e.* a limited knowledge on the “true” simulator. This source of uncertainty can be taken
26 into account by treating the GP model from a Bayesian perspective. This provides the full
27 posterior probability distribution associated with the sensitivity measures, which can be
28 summarized by a confidence interval to outline the regions where the GP model is “unsure”.
29 We show that this methodology is able to provide useful guidelines for the practical decision-
30 making process and suggest further site investigations.

31

32 Keywords: Landslide model; computationally demanding code; global sensitivity analysis;
33 Sobol’ indices; meta-model; Gaussian Process.

34 **1 Introduction**

35 Landslides are very complex phenomena controlled by a range of processes. Geological
36 history, lithology and structure, slope relief and shape, weather and climate, seismicity and
37 human activity can be identified as the main causative factors (Crosta and Clague, 2009). The
38 associated risk to communities can be high (*e.g.* Evans *et al.*, 2002) and thus, predicting
39 landslide behaviour is a major concern.

40 Due to the recent advances in computer modelling (*e.g.* in processor performance) and in
41 particular in the finite element method (*e.g.* van den Ham *et al.*, 2009), numerical models are
42 commonly used in practice to get a better understanding of the landslide behaviour and to
43 predict its evolution. The main drawback of such models is the high number of input factors
44 required for analysis. Global sensitivity analysis of complex numerical models can then be

45 used to determine: (1) which input factors contribute the most to the output variability (within
46 the “factors’ prioritisation setting” as described by Saltelli *et al.*, 2008); (2) which input
47 factors interact with each other; and (3) which input factors are insignificant and can be
48 eliminated to “simplify” the model (within a “factors’ fixing setting”, Saltelli *et al.*, 2008).
49 Such an analysis is useful in identifying which input factors require further investigations to
50 reduce uncertainties in the computer code results, hence providing guidelines for risk
51 management (Saltelli, 2002b).

52 Among the existing sensitivity methods, variance-based methods have proved to be effective
53 (Saltelli *et al.*, 2000). In this article, we focus on the method of Sobol’ indices (Sobol’, 1993;
54 Archer *et al.*, 1997; Sobol’ and Kucherenko, 2005). Unlike traditional linear or rank
55 regression-based methods, these indices allow representing the sensitivity of a general model
56 without assuming any kind of linearity or monotonicity in the model (Saltelli and Sobol’,
57 1995). In practice, the computation of Sobol’ indices uses a Monte Carlo sampling strategy.
58 An example of application in the field of landslide modelling with applications of moderate
59 complexity is provided by Hamm *et al.* (2006). Such an approach, however, appears hardly
60 applicable for more computationally demanding models, as it requires a large number of
61 computer code evaluations. For instance, the study of Hamm *et al.*, (2006) required ten
62 thousand model realisations, corresponding to about 20 hours of computation time (on a 2
63 GHz Pentium 4 PC). The same sensitivity analysis would require 208 days using a model that
64 takes 30 minutes and 2500 days using a model that takes 6 hours to compute.

65 To overcome this difficulty, a first solution is to use a distributed parallel computing
66 methodology, thus requiring an appropriate grid computing architecture and the optimization
67 of computing resources (*e.g.* Dupros *et al.*, 2006; Boulahya *et al.*, 2007).

68 In this paper, an alternative is proposed using a limited number of computer code runs (also
69 named “simulator”, O’ Hagan, 2006), which consists in replacing (*i.e.* approximating) the

70 simulator by a surrogate model with low computation time, also named a “meta-model”, to
71 compute the Sobol’ indices (*i.e.* the sensitivity measures). Various “meta-models” exist (e.g.
72 linear regression, nearest neighbour method, Multivariate Adaptive Regression Spline,
73 neural network and Gaussian Process); see, for example, Storlie *et al.*, 2009 for a recent
74 review.

75 The meta-model uses a limited number of simulator runs, *i.e.* input-output pairs
76 (corresponding to the training sample), to infer the values of the complex simulator output
77 given a yet-unseen input configuration. Such an approximation introduces a source of
78 uncertainty referred to as “code uncertainty” associated with the meta-model (O’ Hagan,
79 2006), so that the sensitivity measures computed with the meta-model are “uncertain”.

80 In the present article, we choose to solve the described problem of approximation (and of
81 inference) under the Bayesian formalism treating the simulator as an “unknown” function in
82 the sense that the simulator output for any yet-unseen input configuration is unknown until the
83 simulator is actually run for the considered configuration (Oakley and O’ Hagan, 2004). We
84 choose to use the concept of an emulator corresponding to a statistical approximation so that a
85 prior probability distribution is assigned to the simulator outputs and updated according to the
86 usual Bayesian paradigm given the training sample. This approach returns not only the most
87 likely value for the output given any input configuration, but also an entire probability
88 distribution (O’ Hagan, 2006). This distribution can be used to estimate a level of confidence
89 when the predictive quality of the meta-model is not high due to a small training data (see, for
90 instance, Marrel *et al.*, 2008 and 2009, Storlie *et al.*, 2009). A Gaussian Process (GP) is
91 chosen as the prior model for the simulator. It has been widely used when designing computer
92 experiments (Sacks *et al.*, 1989; Kennedy and O’Hagan, 2001; Santner *et al.*, 2003).

93 In the first section, the Sobol’ decomposition method is described in the general framework of
94 the variance-based global sensitivity approach (Saltelli *et al.*, 2008).

95 Then, the GP model used as a meta-model of the computationally intensive simulator is
 96 described in the framework of the stochastic processes for computer code experiments under
 97 the Bayesian regression formalism. The methodology for computing the Sobol' indices using
 98 the GP model is described and illustrated in two applications. The first application is a simple
 99 analytical model based on "infinite slope analysis" (Hansen, 1984). This allows us to compare
 100 the sensitivity measures computed using the "true" model with those computed using the GP
 101 model. Finally, the application of this methodology to a La Frasse (Switzerland) landslide
 102 model (Laloui *et al.*, 2004) is presented and we show how to use the sensitivity measures to
 103 guide the decision-making process for further site investigations.

104 **2 Global sensitivity analysis by the Sobol' decomposition method**

105 **2.1 Introduction on the variance-based sensitivity analysis**

106 Consider the simulator g and the scalar output y determined from a vector of n input factors
 107 $\mathbf{x} = \{x_i\}_{i=1,\dots,n}$ so that $y = g(\mathbf{x})$.

108 Considering the n -dimensional vector as a random vector of independent random variable X_i ,
 109 then the output Y is also a random variable (as a function of a random vector). A variance-
 110 based sensitivity analysis aims at determining the part of the total unconditional variance V_Y
 111 of the output Y resulting from each input random variable X_i . The total variance V_Y can be
 112 expressed as follows (Saltelli *et al.*, 2000 & 2008):

113

$$114 \quad V_Y = \sum_i V_i + \sum_{i < j} V_{ij} + \sum_{i < j < l} V_{ijl} + \dots + V_{ij\dots n} \quad (1)$$

115

116 where the partial variance V_i and $V_{ij\dots n}$ read:

117

118 $V_i = Var[E[Y|X_i = x_i]]$ (2)
 119 $V_{ij\dots n} = Var[E[Y|X_i = x_i, X_j = x_j, \dots, X_n = x_n]] - V_i - V_j - \dots - V_n$

119

120 with $E[Y|X_i = x_i]$, the expectation of Y given that the i^{th} input factor X_i has a fixed value x_i

121 and $E[Y|X_i = x_i, X_j = x_j, \dots, X_n = x_n]$ the conditional expectation of Y given that the i^{th} input

122 factor X_i has a fixed value x_i , the j^{th} input factor X_j has a fixed value x_j , ...etc.

123 The variance of the conditional expectation V_i represents the first order effect of the input

124 factor X_i taken alone, whereas the higher order indices account for possible mixed influence

125 of various input factors.

126

127 **2.2 The Sobol' decomposition method**

128 **2.2.1 Presentation**

129 To determine the partial variances of Y , Sobol' (1993) proposes the following decomposition

130 of g into summands of increasing dimension provided that g is integrable:

131

132 $g(\mathbf{x}) = g_0 + \sum_{i=1}^n g_i(x_i) + \sum_{i < j} g_{ij}(x_i, x_j) + \dots + g_{1,\dots,n}(x_1, \dots, x_n)$ (3)

133

134 where g_0 corresponds to the mean constant value of the function g and each term can be

135 evaluated through multidimensional integrals as follows:

136

137 $g_0 = E[Y] = \int_0^1 \dots \int_0^1 g(\mathbf{x}) d\mathbf{x} = \int_0^1 \dots \int_0^1 g(x_1, \dots, x_n) dx_1 \dots dx_n$ (4)

138 $g_i(x_i) = E[Y|X_i = x_i] - g_0 = \int_0^1 \dots \int_0^1 g(\mathbf{x}) d\mathbf{x}_{-i} - g_0$ (5)

$$g_{ij}(x_i, x_j) = E[Y|X_i = x_i, X_j = x_j] - g_0 - g_i - g_j = \int_0^1 \dots \int_0^1 g(\mathbf{x}) d\mathbf{x}_{-(ij)} - g_0 - g_i(x_i) - g_j(x_j) \quad (6)$$

140

141 with $d\mathbf{x}_{-i}$ denotes the integration over all input factors except x_i and $d\mathbf{x}_{-(ij)}$, the integration

142 over all input factors except both x_i and x_j . Similar formulae can be obtained for higher order

143 terms.

144

145 The total variance V_Y can then be expressed as:

146

$$V_Y = \int_0^1 \dots \int_0^1 g(x_1, \dots, x_n)^2 dx_1, \dots, dx_n - g_0^2 \quad (7)$$

148

149 while the partial variances read as follows:

150

$$V_{i_1 \dots i_s} = V[g_{i_1 \dots i_s}] = \int_0^1 \dots \int_0^1 g_{i_1 \dots i_s}(x_{i_1}, \dots, x_{i_s})^2 dx_{i_1} \dots dx_{i_s} \quad (8)$$

152 with $1 \leq i_1 < \dots < i_s \leq n$ and $s = 1, \dots, n$.

153 2.2.2 Definition of the Sobol' indices

154 The Sobol' indices $S_{i_1 \dots i_s}$ describe which amount of the total variance is due to the

155 uncertainties of input factors in the set $\{i_1, \dots, i_s\}$ and is expressed as the ratio between $V_{i_1 \dots i_s}$

156 and V_Y , respectively the partial and total variances .

157 The first-order sensitivity index S_i for input factor X_i is expressed as follows:

158

$$S_i = \frac{V_i}{V_Y} \quad (9)$$

160 The sensitivity measure S_i is referred to as “the main effect of X_i ” and can be interpreted as
161 the expected reduction in the total variance of the output Y (*i.e.* representing the uncertainty in
162 Y) if the true value of the input factor X_i was known. This index provides a measure of
163 importance useful to rank the input factors (Saltelli *et al.*, 2000 & 2008).

164 The main effect and the higher order Sobol’ indices satisfy the following property:

165

$$166 \quad \sum_{i=1}^n S_i + \sum_{i < j} S_{ij} + \dots + S_{ij\dots n} = 1 \quad (10)$$

167

168 Defining S_{-i} as the sum of all the terms $S_{i_1\dots i_s}$ but the i^{th} index, the total effect index S_{Ti} of X_i
169 is defined as the total contribution of the i^{th} input factor to the total variance. It reads as
170 follows (using eq. 10):

171

$$172 \quad S_{Ti} = 1 - S_{-i} \quad (11)$$

173

174 $S_{Ti} = 0$ means that the input factor X_i has no effect. Thus, it can be fixed at any value over its
175 uncertainty range (Saltelli *et al.*, 2008).

176 As the total number of sensitivity indices reaches $2^n - 1$ (Saltelli *et al.*, 2000), hence
177 representing a high computational cost, the sensitivity analysis is generally limited, in practice,
178 to the pair of indicators corresponding to the main effect S_i and to the total effect S_{Ti} of X_i
179 (Saltelli *et al.*, 2008).

180 2.2.3 Numerical implementation

181 The evaluation of the Sobol’ indices can be carried out through a Monte Carlo sampling
182 strategy (Saltelli *et al.*, 2000), which remains an approximation of the true value of the

183 sensitivity indices. Thus, the quality of the approximation directly depends on the sample size.

184 Let us consider m sampled elements $\{\mathbf{x}^{(j)}\}_{j=1,\dots,m}$ in the n -dimensional space of input factors:

185

$$\begin{cases} \hat{g}_0 \approx \frac{1}{m} \sum_{j=1}^m g(\mathbf{x}^{(j)}) \\ \hat{V}_Y \approx \frac{1}{m} \sum_{j=1}^m g^2(\mathbf{x}^{(j)}) - \hat{g}_0^2 \\ \hat{V}_i \approx \frac{1}{m} \sum_{j=1}^m g(\mathbf{x}_{S_a}^{-(ij)}, \mathbf{x}_{S_a}^{(ij)}) \cdot g(\mathbf{x}_{S_b}^{-(ij)}, \mathbf{x}_{S_b}^{(ij)}) - \hat{g}_0^2 \end{cases} \quad (12)$$

187

188 where $\mathbf{x}_{-ij} = [x_{1j}, \dots, x_{(i-1)j}, x_{(i+1)j}, \dots, x_{nj}]$ and the subscripts S_a and S_b in eq. (12) indicate that two

189 sampling data matrices are being used.

190 The main and total effects can be estimated using the sampling strategy proposed by Saltelli

191 (2002a) at a computation time cost of $m \times (n+2)$ model evaluations. Additional computational

192 efficiency can be achieved by making best use of sampling designs, for instance Sobol' quasi-

193 random sequences, and estimators, for instance Jansen's estimator (Saltelli *et al.*, 2010).

194 However, the computational effort for simulators with computation time ranging from several

195 hours to several days may still be high and the present work focuses on a strategy based on

196 Gaussian Process meta-modelling to reduce this effort.

197 **3 Gaussian Process (GP) modelling**

198 **3.1 Description of the stochastic process framework**

199 First, the deterministic response $y(\mathbf{x})$ of the simulator is treated as a realization of a random

200 variable $Y(\mathbf{x})$, which can be decomposed into a deterministic function f , which represents the

201 mean (*i.e.* expectation of Y), and a stochastic function Z as proposed by Sacks *et al.* (1989)

202 and reads as follows:

203

$$204 \quad Y(\mathbf{x}) = f(\mathbf{x}) + Z(\mathbf{x}) \quad (13)$$

205

206 Note that the case of multiple outputs is beyond the scope of the present article and the
207 interested reader is advised to refer to Le and Zidek (2006).

208 Without prior information between the modelling inputs and outputs, f is chosen as a
209 multivariate linear regression model (Martin and Simpson, 2005) so that:

210

$$211 \quad f(\mathbf{x}) = \mathbf{F}(\mathbf{x}) \cdot \mathbf{B} = \sum_{i=0}^n b_i \cdot f_i(\mathbf{x}) \quad (14)$$

212

213 where $\mathbf{B} = \{b_i\}_{i=0, \dots, n}$ is the regression parameter vector; $\mathbf{F}(\mathbf{x}) = [f_0(\mathbf{x}), f_1(\mathbf{x}), \dots, f_n(\mathbf{x})]$ is the
214 corresponding regression vector with f_i ($i=0, 1, \dots, n$), the basis functions. Assuming linearity
215 for the mean f , we have: $f_0(\mathbf{x})=1$ and $f_i(\mathbf{x})=x_i$ for $i=1, \dots, n$.

216 The stochastic part Z can be seen as a confidence measure on the model output mean. It
217 represents a zero mean random process, characterized by its $n \times n$ covariance matrix Σ_S so that
218 an element at the j^{th} row and k^{th} column of Σ_S is expressed as:

219

$$220 \quad \Sigma_{S,j,k} = \text{Cov}[Z(\mathbf{x}^{(j)}), Z(\mathbf{x}^{(k)})]_{j,k=1 \dots N_S} = \sigma^2 \cdot \rho(\mathbf{x}^{(j)}, \mathbf{x}^{(k)})_{j,k=1 \dots N_S} \quad (15)$$

221

222 where $\mathbf{x}^{(j)}$ is the j^{th} input vector (with $j=1 \dots N_S$) and $\text{Var}[Z]=\sigma^2$. The correlation function ρ
223 provides the interpolation and spatial correlation properties.

224 Treating the simulator g within the stochastic process framework allows the deterministic part
225 f to account for the global behaviour of g , whereas the correlation terms allow the meta-model
226 to “locally” interpolate the known data by introducing a strong correlation in the

227 neighbourhood of these points. Complex input-output behaviours are hence better represented
228 (e.g. see Langewisch and Apostolakis, 2010).

229 **3.2 Description of the Gaussian correlation model**

230 Various authors (e.g. Stein, 1999; Le and Zidek, 2006) have discussed different types of
231 correlations functions. For our purposes, the study is restricted to the Gaussian correlation
232 model so that the value of the correlation matrix only depends on the normalised distance
233 between the input vectors $\mathbf{x}^{(j)}$ and $\mathbf{x}^{(k)}$. Assuming the correlation model is invariant to any
234 translation in the input space (e.g. Rasmussen and Williams, 2006), the Gaussian correlation
235 function reads as follows:

236

$$237 \quad \rho(\mathbf{x}^{(j)}, \mathbf{x}^{(k)}) = \prod_{i=1}^n \rho_i(x_i^{(j)}, x_i^{(k)}) = \exp \left\{ - \sum_{i=1}^n \frac{\|x_i^{(j)} - x_i^{(k)}\|^2}{\theta_i} \right\} \quad (16)$$

238

239 where $\boldsymbol{\theta} = \{\theta_i\}_{i=1, \dots, n}$ are the correlation lengths, also referred to as “hyper-parameters” (e.g. see
240 Rasmussen and Williams, 2006). θ_i parameter quantifies the rate at which the output varies as
241 i^{th} input factor is changed.

242 Note that in case of data measurements errors or non deterministic computer code, a constant
243 regularization term referred to as “nugget effect” may be defined, hence introducing a white
244 noise.

245 **3.3 Principle and prediction under the Bayesian formalism**

246 In this paper, we focus on the stationary GP model which fits the stochastic framework and
247 has been broadly used in designing computer experiments (Sacks *et al.*, 1989; Kennedy and
248 O’Hagan, 2001; Santner *et al.*, 2003).

249 Let us define the training sample as the N_S training data pairs $\{\mathbf{X}_S, \mathbf{Y}_S\}$, which represent a
 250 mapping between the spaces of input factors $\mathbf{X}_S = \{\mathbf{x}^{(j)}\}_{j=1\dots N_S}$, with $\mathbf{x}^{(j)} = \{x_i^{(j)}\}_{i=1,\dots,n}$, and the
 251 outputs $\mathbf{Y}_S = \{y^{(j)}\}_{j=1\dots N_S}$, obtained through the N_S selected simulator runs so that
 252 $y^{(j)} = \{f(\mathbf{x}^{(j)})\}_{j=1,\dots,N_S}$.

253 In a first step, constructing the GP model implies considering the simulator output Y as a
 254 random variable, which is assumed to follow a multivariate Gaussian distribution (denoted G)
 255 for any random vector of input factors \mathbf{X} . This assumption represents our prior belief on the
 256 simulator. Using the training sample $\{\mathbf{X}_S, \mathbf{Y}_S\}$, the Bayes theorem is used to refine the
 257 mentioned prior information in order to yield the posterior distribution of the output, known
 258 as the “emulator” (e.g. O’ Hagan, 2006). This latter not only provides an expected value for
 259 any “yet-unseen” input configuration, but it also gives an entire posterior distribution given
 260 the observed data.

261 Formally, the probability $p(\mathbf{Y}_S|\mathbf{X}_S)$ of obtaining \mathbf{Y}_S given \mathbf{X}_S , is expressed under the GP
 262 assumption, as follows:

$$263 \quad p(\mathbf{Y}_S|\mathbf{X}_S) \sim G(F(\mathbf{X}_S), \mathbf{B}, \Sigma_S) \quad (17)$$

264 Considering a new vector $\mathbf{x}^{(N_S+1)}$ of input factors and the associated output $Y^{(N_S+1)} = Y(\mathbf{x}^{(N_S+1)})$,
 265 the joint probability distribution of the random variables $(\mathbf{Y}_S, Y^{(N_S+1)})$ reads as follows:

$$266 \quad p(\mathbf{Y}_S, Y^{(N_S+1)}|\mathbf{X}_S, \mathbf{B}, \sigma, \boldsymbol{\theta}) \sim G\left(\begin{bmatrix} \mathbf{F}_S \\ F(\mathbf{x}^{(N_S+1)}) \end{bmatrix} \cdot \mathbf{B}, \begin{bmatrix} \Sigma_S & k(\mathbf{x}^{(N_S+1)}) \\ k(\mathbf{x}^{(N_S+1)})^T & \sigma^2 \end{bmatrix}\right) \quad (18a)$$

267 where $\mathbf{F}_S = [F(\mathbf{x}^{(1)}), F(\mathbf{x}^{(2)}), \dots, F(\mathbf{x}^{(N_S)})]$ corresponds to the regression matrix.

268 $k(\mathbf{x}^{(N_S+1)})$ represents the vector of correlation functions between each of the N_S training input
 269 vectors $\mathbf{x}^{(j)}$ and the new element $\mathbf{x}^{(N_S+1)}$. It can be written as:

270

271 $k(\mathbf{x}^{(N_s+1)}) = \sigma^2 \cdot [\rho(\mathbf{x}^{(1)}, \mathbf{x}^{(N_s+1)}), \dots, \rho(\mathbf{x}^{(N_s)}, \mathbf{x}^{(N_s+1)})]^T$ (17b)

272

273 Within the Bayesian framework, the posterior distribution of the computed output random

274 variable $Y^{(N_s+1)}$ is conditioned on the “observed” (i.e. actually calculated) values

275 corresponding to the training sample $\{\mathbf{X}_S, \mathbf{Y}_S\}$, given the new element $\mathbf{x}^{(N_s+1)}$, and follows a

276 multivariate Gaussian distribution (Von Mises, 1964):

277

278
$$p(Y^{(N_s+1)} | \mathbf{Y}_S, \mathbf{X}_S, \mathbf{B}, \sigma, \boldsymbol{\theta}) = \frac{p(\mathbf{Y}_S, Y^{(N_s+1)} | \mathbf{X}_S, \mathbf{B}, \sigma, \boldsymbol{\theta})}{p(\mathbf{Y}_S | \mathbf{X}_S, \mathbf{B}, \sigma, \boldsymbol{\theta})} \quad (18a)$$

$$\sim G(E[Y^{(N_s+1)} | \mathbf{Y}_S, \mathbf{X}_S, \mathbf{B}, \sigma, \boldsymbol{\theta}], \text{Var}[Y^{(N_s+1)} | \mathbf{Y}_S, \mathbf{X}_S, \mathbf{B}, \sigma, \boldsymbol{\theta}])$$

279 where:

280 $E[Y^{(N_s+1)} | \mathbf{Y}_S, \mathbf{X}_S, \mathbf{B}, \sigma, \boldsymbol{\theta}] = F(\mathbf{x}^{(N_s+1)}) \cdot \mathbf{B} + k(\mathbf{x}^{(N_s+1)})^T \cdot \boldsymbol{\Sigma}_S^{-1} \cdot (\mathbf{Y}_S - \mathbf{F}_S \cdot \mathbf{B})$ (18b)

281 and:

282 $\text{Var}[Y^{(N_s+1)} | \mathbf{Y}_S, \mathbf{X}_S, \mathbf{B}, \sigma, \boldsymbol{\theta}] = \sigma^2 - k(\mathbf{x}^{(N_s+1)})^T \cdot \boldsymbol{\Sigma}_S^{-1} \cdot k(\mathbf{x}^{(N_s+1)})$ (18c)

283

284 with $\boldsymbol{\Sigma}_S^{-1}$, the inverse of the covariance matrix associated to the training input data \mathbf{X}_S .

285 The conditional mean of eq. (18b) is used as a predictor and the conditional variance in eq.

286 (18c) corresponds to the mean square error of the predictor term. Provided that the new

287 candidate $\mathbf{x}^{(N_s+1)}$ is far away from the training input data \mathbf{X}_S , the term

288 $k(\mathbf{x}^{(N_s+1)})^T \cdot \boldsymbol{\Sigma}_S^{-1} \cdot k(\mathbf{x}^{(N_s+1)})$ will be small so that the predicted variance will be large.

289 In a more general manner, if we consider two new test candidates \mathbf{u} and \mathbf{v} , the general

290 expression of the conditional GP model can be written as:

291

292 $(Y|Y_S, X_S, \mathbf{B}, \sigma, \boldsymbol{\theta}) \sim GP(E[Y(\mathbf{u})|Y_S, X_S, \mathbf{B}, \sigma, \boldsymbol{\theta}], \text{Cov}[Y(\mathbf{u}), Y(\mathbf{v})|Y_S, X_S, \mathbf{B}, \sigma, \boldsymbol{\theta}])$ (19)

293

294 The conditional mean is used as a predictor and is expressed as in eq. (18b).

295 The conditional covariance provides the confidence on the prediction and reads as follows:

296

297 $\text{Cov}[Y(\mathbf{u}), Y(\mathbf{v})|Y_S, X_S, \mathbf{B}, \sigma, \boldsymbol{\theta}] = \sigma^2 \rho(\mathbf{u}, \mathbf{v}) - k(\mathbf{u})^T \cdot \Sigma_S^{-1} \cdot k(\mathbf{v})$ (20)

298

299 The regions of the input space, where few data are available will be underlined by higher
300 variance.

301 The main difficulty in constructing a conditional GP model given a training sample resides in

302 determining the parameters corresponding to the regression coefficient vector \mathbf{B} , the hyper-

303 parameters $\boldsymbol{\theta}$ and the variance σ^2 . A first approach consists in estimating them as solutions of

304 the optimization problem using the maximum likelihood method, *e.g.* implemented in the

305 GEM-SA software (O'Hagan, 2006) and the MATLAB toolbox DACE (Lophaven, 2002).

306 However, the optimisation algorithms used for the parameters identification may show

307 limitations, especially in case of high dimension problem (*e.g.* see Marrel *et al.*, 2008).

308 Besides, such an approach may underestimate the variance in the predictions of new

309 observations (Cressie, 1993).

310 In this paper, an approach based on the Bayesian framework (*e.g.* Rasmussen 1996) is chosen

311 so that the hyper-parameters are given prior distributions $p(\boldsymbol{\theta})$. In the Bayesian framework, the

312 Markov Chain Monte Carlo methods (Gilks *et al.*, 1996) are used to integrate over the

313 posterior distribution $p(\boldsymbol{\theta}|X_S, Y_S)$ associated with the GP parameters *i.e.* the training sample is

314 used to update the GP parameters. The posterior distribution of the hyper-parameters will be

315 hence concentrated on values that are consistent with actually observed data. This procedure

316 is implemented in the package named "TGP" of the "R" software ("R" Development Core

317 Team, 2009) by Gramacy and co-workers (Gramacy, 2007; Gramacy and Taddy, 2010).
318 Further theoretical details can be found in Gelman *et al.* (1995) and Gramacy and Herbert
319 (2009). Though computationally more intensive (Storlie *et al.*, 2009), this approach presents
320 the attractive feature to incorporate the uncertainty related to the construction of the GP model,
321 so that the level of confidence associated with the “meta-model”-computed sensitivity
322 measures also takes this source of uncertainty into account (see step 4 of section 3.4.).

323 **3.4 A “GP-based” methodology for sensitivity analysis**

324 In this section, we describe the methodology to compute the sensitivity measures (*i.e.* the
325 Sobol’ indices) using a GP model as a surrogate model of the computationally intensive
326 simulator.

327 **3.4.1 Step 1: representation of the input factor uncertainty**

328 The first step is to characterize and mathematically represent the uncertainty (range and form
329 of the probability distribution) on each of the input factors. This representation can have a
330 strong influence on the final sensitivity results, hence on risk management decision making
331 (Saltelli, 2002b).

332 Representing the uncertainty through empirical probability distributions requires a large
333 amount of data (laboratory or *in situ* measurements), which may not be practical in many
334 situations. Thus, knowledge on the range of uncertainty is commonly evaluated either based
335 on physical reasoning, on analogies with similar cases or simply from expert opinions,
336 whereas the mathematical representation of the probability distribution may either be
337 theoretically known or assumed. In a situation where “sparse, vague and incomplete” data are
338 available, a common approach consists in assigning a uniform probability distribution based
339 on the “maximum entropy” approach (Gzyl, 1995).

340 3.4.2 Step 2: setting training data

341 The objective then is to run the simulator for a limited number of times N_S in order to create a
342 mapping between the input factor and the computer code output domain. The number N_S
343 should be defined as a compromise between the minimization of the computation time cost
344 and the maximization of the input factor domain exploration (directly linked with the
345 accuracy and reliability of the GP model, see step 3).

346 In this view, we propose to use the Latin hypercube sampling method (McKay *et al.*, 1979) in
347 combination with the “maxi-min” space-filling design criterion (Koehler and Owen, 1996).
348 More sophisticated strategies exist mainly based on sequentially adaptive design of
349 experiments adding new training candidates where the predictive uncertainty is high (*e.g.*
350 Gramacy and Herbert, 2009). The use of such approaches is beyond the scope of this paper.

351 3.4.3 Step 3: constructing the GP model

352 Using the GP model instead of the simulator introduces an additional source of uncertainty
353 referred to as “code uncertainty” (O’ Hagan, 2006). In the regions where the true simulator is
354 not evaluated, we are uncertain about what the “true” simulator would introduce. This sort of
355 uncertainty can be reduced by increasing our knowledge of the true simulator, *i.e.* by
356 increasing the training sample size.

357 Except when a “nugget” effect is included, the GP model is an exact interpolator, so that
358 residuals of the training data cannot be directly used to validate the approximation (Marrel *et*
359 *al.*, 2008). The key aspect for validating the “statistical” approximation is to estimate the
360 expected level of fit (*i.e.* predictive quality) of the GP model to a data set that is independent
361 of the data (*i.e.* “yet-unseen” data) that were used to train the GP model.

362 As additional simulator runs are costly, using a test sample of new data might be impractical
363 and cross-validation procedures such as the “ k -fold” cross-validation technique (Hastie, 2002)
364 should be used. In this cross-validation procedure, the initial training sample is randomly

365 partitioned into k subsets. In a first step, a single subset is used as the validation data for
366 testing the GP model, and the remaining $k-1$ subsets are used as training data for the
367 construction of the GP model. For each step, the k validation data are estimated and the
368 coefficient of determination R^2 for the procedure is computed as follows:

$$369 \quad R^2(\mathbf{y}, \hat{\mathbf{y}}) = 1 - \frac{\sum_{i=1}^k (\hat{y}_i - y_i)}{\sum_{i=1}^k (y_i - \bar{y})} \quad (21)$$

370 where \mathbf{y} represents the vector of observations in the validation set; \bar{y} is the mean of the
371 corresponding sample and $\hat{\mathbf{y}}$, the vector of predicted values using the GP model.

372 The coefficient R^2 provides a metric of the predictive quality so that a value close to 100 %
373 indicates that the GP model is successful in matching the validation data. A typical threshold
374 of 80 % is commonly used to qualify the predictive quality as “satisfactory” (e.g. Marrel et
375 al., 2008).

376 The cross-validation process is then repeated k times using each of the k subsets as validation
377 samples. For small training sets, the cross validation procedure with $k=1$ is usually used
378 corresponding to the so-called “leave-one-out” cross validation procedure.

379 **3.4.4 Step 4: estimating the sensitivity measures**

380 The most likely value $\mu(S_i)$ for the sensitivity measures is computed using the conditional
381 mean of the GP model in eq. 18b. Additional useful information for risk management
382 purposes is the level of confidence (or accuracy) related to the sensitivity analysis based on
383 the GP model. A confidence interval $CI(S_i)$ can be defined with bounds corresponding to the
384 5% and to the 95 % quantile of the full posterior distribution of the sensitivity measures. This
385 confidence interval both summarizes the “code uncertainty” associated with the meta-model
386 (O’ Hagan, 2006) and the uncertainty on the estimation of the GP model parameters (see
387 section 3.3).

388 4 Illustrative analytical model

389 In this section, we consider the infinite slope analytical model (*e.g.* Hansen, 1984) in order to
390 illustrate the methodology described in section 3.4. This model is of course not a
391 computationally demanding function, but we imagine it as representing a calculation that may
392 take several minutes or even hours of computation to evaluate. Besides, using this analytical
393 model also allows us to compare the results of the sensitivity analysis using the “true” model
394 with those using the GP model.

395 4.1 Description of the analytical model

396 The stability of the infinite slope model as depicted in Fig. 1 is evaluated by deriving the
397 factor of safety FS , which corresponds to the ratio between the resisting and the driving forces
398 acting on the slope (eq. 22):

$$400 \quad FS = \frac{C + (\gamma - m \cdot \gamma_w) \cdot z \cdot \cos(\theta) \cdot \cos(\theta) \cdot \tan(\phi)}{\gamma \cdot z \cdot \sin(\theta) \cdot \cos(\theta)} \quad (22)$$

401

402 *[Fig. 1 about here]*

403

404 where C is the cohesion of the soil material; ϕ , the friction angle; θ , the slope angle; γ , the
405 soil unit weight; γ_w , the water unit weight; z , the thickness of slope material above the slip
406 plane; and m , the ratio between thickness of surficial saturated slope material and z . If FS is
407 lower than 1.0 the potential for failure is high.

408 **4.2 Sensitivity analysis of the analytical model**

409 For illustrative purposes, we only considered the thickness z and the slope angle θ as
410 uncertain input factors. The other input factors were assumed fixed: $C=10$ kPa, $\phi=25^\circ$,
411 $\gamma=22$ kN.m⁻³, $\gamma_w=9.81$ kN.m⁻³ and $m=90$ %.
412 The objective was to identify whether z or θ contributes the most to the FS variability within a
413 “factors’ prioritisation setting”. It is assumed that very sparse data are available to
414 characterize the uncertainty on these input factors so that z uniformly varies between 5 and 25
415 m and θ uniformly varies between 25° and 35° (step 1). We generated two different training
416 samples of respectively 6 and 20 training data of the form $\{z ; \theta ; FS\}$, using the Latin
417 hypercube sampling approach (step 2) and for each training sample, a GP model was
418 constructed.

419

420 *[Fig. 2 about here]*

421

422 Fig. 2 (top) shows the comparison between the values of FS obtained from direct simulations
423 on a grid of 10×10 in the input factor domain $[5 ; 25] \times [25 ; 35]$ (straight line) and from the
424 prediction on the same grid using the GP model (dashed line) for both training samples (Fig.
425 2, left for 6 training data and right for 20 training data). The coefficient of determination or
426 goodness of fit (eq. 22) estimated for both GP models was equal to 90.9 % for the first
427 training sample and to 98.8 % for the second one, hence showing a very good match for both
428 meta-models. The quality of the approximation was then estimated through a “leave-one-out”
429 cross validation procedure (step 3): we obtained a coefficient of determination of 96.2 % for
430 the first sample and 99.7 % for the second one, hence indicating a “high” predictive quality.
431 The estimated FS using both GP models (Fig. 2, middle) were compared to the “true”
432 observed FS . The closer the dots to the straight black line, the better the approximation.

433 The results for the computation of the main effects required (step 4)
434 $m \times (n+2) = 2500 \times (2+2) = 10000$ model evaluations using the sampling strategy of Saltelli
435 (2002a). The most likely of the main effects calculated with both GP models (blue dots in Fig.
436 2, bottom) were compared to the main effects obtained from direct simulations (red dots on
437 Fig. 2, bottom) by means of the R package “sensitivity” and the function referred to as
438 Sobol2002 (available at <http://cran.r-project.org/web/packages/sensitivity/index.html>). These
439 results are summarized in Table 1.

440

441 *[Table 1 about here]*

442

443 We see that differences are larger for the GP model constructed with the lower training
444 sample size but, however, the “true” values for the main effects still lie within the confidence
445 interval bounded by the 5 % and the 95 % quantile (black cross-type marker in Fig. 2,
446 bottom). Not surprisingly, increasing the number of training samples (*i.e.* our knowledge of
447 the true function) decreases the range of code uncertainty as well as the differences between
448 the true values and estimates (Table 1).

449 **5 Computationally intensive numerical model**

450 In this section, we present the application of the proposed GP-based sensitivity analysis
451 methodology (section 3.4.) to the landslide finite-element model originally used for the
452 simulation of the La Frasse (Switzerland) landslide during the 1994 crisis period (Laloui *et al.*,
453 2004).

454 **5.1 General description of the landslide model**

455 The La Frasse landslide covers a total area of roughly 1000×1000 m², and represents an
456 average thickness of 80 m in its upper part and 40 m in its lower part. The total volume of the

457 La Frasse landslide reaches 73 million m³. Since 1975, a constant movement has been
458 observed in its upper and central parts, varying between 10.10⁻² and 15.10⁻² m per year
459 (Noverraz and Bonnard, 1988; Bonnard *et al.*, 1995). The evolution of the groundwater table
460 is considered to be at the origin of the sliding and the instabilities were mainly observed
461 during the 1994 crisis (over a period of nearly 300 days). Therefore, in order to assess the
462 effect of the hydraulic regime on the geomechanical behaviour of the landslide, finite-element
463 simulations considering a 2D cross-section through the centre of the landslide were performed
464 by Laloui *et al.* (2004) using the finite element program GEFDYN (Aubry *et al.*, 1986).
465 The model is composed of 1694 nodes, 1530 quadrangular elements, and six soil layers
466 derived from the geotechnical investigations. Fig. 3 gives an overview of the model, as well
467 as the boundary conditions used for analysis. Instabilities observed in 1994 were triggered by
468 pore pressure changes occurring at the base of the slide (see Laloui *et al.*, 2004 for further
469 details).

470

471 *[Fig. 3 about here]*

472

473 The general behaviour of the landslide is strongly correlated to the properties of the slip
474 surface. The complex behaviour of the slip surface material was modelled using the Hujieux
475 elastoplastic multi-mechanism constitutive model (Aubry *et al.*, 1982; Hujieux, 1985; Lopez-
476 Caballero *et al.*, 2007; Lopez-Caballero and Modaressi Farahmand-Razavi, 2008) and the
477 Mohr-Coulomb criterion was assumed for the other soil materials.

478 The Hujieux constitutive model permits coverage of a large range of deformation and takes
479 into account: (1) the influence of confinement and stress path on the moduli; (2) the effects of
480 over-consolidation; and (3) the influence of the void ratio. It can be used for granular as well
481 as clayey soil behaviours and it is based on a Coulomb type failure criterion and the critical

482 state concept. The volumetric and deviatoric hardening regimes implemented in the Hujeux
483 model lead to a dependence on the consolidation pressure as in the Cam-Clay family models,
484 and to the evolution of the plastic yield surface with the deviatoric and volumetric plastic
485 strains. Moreover, the model accounts for dilatancy/contractance of soils and non-associated
486 flowing behaviour with evolution of the plastic strain rate through a Roscoe-type dilatancy
487 rule.

488 As outlined by Laloui *et al.* (2004), the main parameters for the slip surface materials are: (1)
489 the bulk (K) and shear (G) elastic modules, which are assumed to depend on the mean
490 effective stress through a power-type law of exponent n_e ; (2) the critical state and plasticity
491 parameters, essentially the friction angle ϕ at perfect plasticity, the plastic compressibility β ;
492 and (3) the dilatancy angle Ψ , appearing in the flow rule and defining the limit between soil
493 dilatancy and contractance.

494 Note that these parameters can be directly measured from either *in situ* or laboratory test
495 results (Lopez-Caballero *et al.*, 2007; Lopez-Caballero and Modaressi Farahmand-Razavi,
496 2008). The other Hujeux law parameters, appearing in the flow rule, the hardening and the
497 threshold domains definition are categorized as “not-directly measurable” (Lopez-Caballero
498 and Modaressi Farahmand-Razavi, 2008) and are estimated through numerical calibration
499 techniques between the observed/experimental data and the simulated ones.

500 **5.2 Sensitivity analysis using the GP-based methodology**

501 The sensitivity analysis using the GP-based methodology (see section 3.4.) was carried out to
502 assess the importance of the input factors of the Hujeux constitutive model describing the slip
503 surface behaviour within a “factors’ prioritisation setting”, so that the main effects (first order
504 Sobol’ indices) were used for ranking.

505 The quantity of interest was chosen as the horizontal displacement calculated at two
506 observation points, namely in the upper (observation point 1, Fig. 3), and lower parts of the

507 landslide (observation point 2, Fig. 3). The sensitivity analysis was carried out in a dynamic
508 manner at each step of the 300 days long crisis period (decomposed into a hundred time steps).
509 It was focused on the main measurable parameters of the Hujoux constitutive model (total
510 number of seven input factors), the others being kept constant *i.e.* treated with “no
511 uncertainty”. The properties of the other soil layers were assumed to be constant as well.

512 **5.2.1 Step 1: representation of the input factor uncertainty**

513 In this illustrative study, our objective was to explore the situation where the same “level of
514 uncertainty” is associated with the parameters of Hujoux model: a 25 % variation around the
515 original values identified by Laloui and co-authors (Laloui *et al.*, 2004) was affected to each
516 of the seven input factors (Table 2). Considering no further information on the uncertainty, a
517 uniform probability distribution was assigned to each of these input factors (see section 3.4.1.).

518

519 *[Table 2 about here]*

520

521 **5.2.2 Step 2: setting training data**

522 A total number of 30 input parameter configurations was generated. The resulting horizontal
523 displacements computed over the crisis period are shown on Figure 4 for the observation
524 points 1 and 2. For a given input configuration, a simulator run required ≈ 96 hours on a
525 computer unit (CPU) with a 2.6 GHz dual core processor and 1 GB of RAM. The training
526 sample was generated using a grid computing architecture or computer cluster composed of
527 30 CPU, so that all simulations were performed in parallel.

528

529 *[Fig. 4 about here]*

530

531 5.2.3 Step 3: constructing the GP model

532 At each step of the 1994 crisis period, a GP model was constructed using the 30 training data
533 to approximate the horizontal displacements at the observation points 1 and 2.

534 A “leave-one-out cross-validation” procedure was carried out for each step in order to assess
535 the predictive quality of the GP models. Fig. 5 depicts the temporal evolution of the
536 coefficient of determination R^2 for the cross-validation procedure.

537 During the first half of the crisis period (first 150 days), R^2 decreases over time for both
538 observation points between 99.9 % and ≈ 95 %, hence indicating that the predictive quality is
539 “high” over this period. During the second half of the crisis period, the quality is still
540 satisfactory if we consider observation point 2 (R^2 varying between ≈ 80 % and ≈ 95 %, see
541 Marrel *et al.*, 2008), whereas it can be qualified as “low to moderate” for observation point 1
542 (R^2 steeply decreasing from ≈ 95 % to ≈ 62 %), hence indicating possibly high uncertainty on
543 the GP model.

544

545 *[Fig. 5 about here]*

546

547 5.2.4 Step 4: estimating the sensitivity measures

548 The main effects were calculated using the sampling strategy of Saltelli (2002a), hence
549 requiring $m \times (n+2) = 1000 \times (7+2) = 9000$ GP model evaluations. Preliminary convergence tests
550 were carried out for $m=250, 500, 1000$ and 2000 : they showed that $m=1000$ yields satisfactory
551 convergence of the sensitivity measures to two decimal places (± 0.025).

552 The total computation time of the GP-based sensitivity analysis reached a total of 108 hours
553 (4.5 days), including the generation of the training sample (≈ 4 days), the construction of a GP

554 model at each step of the crisis period (≈ 3 hours) and the cross-validation procedure (≈ 3
555 hours).

556 If the same analysis had been undertaken by direct simulations, the total computation time
557 would have reached $9000/30 \times 96 = 28800$ hours (1200 days) using the same 30 CPU cluster. To
558 achieve a computation time of 108 hours, a computer cluster composed of 8000 CPU would
559 have been required.

560 5.2.5 Analysis of the temporal evolution of the main effects

561 Fig. 6 (top) depicts the temporal evolution of the “first most important” input factor (straight
562 green line) at the observation point 1 in the upper part of the landslide (Fig. 6, left) and at the
563 observation point 2 in the lower part of the landslide (Fig. 6, right). Similarly, Fig. 6 (bottom)
564 provides the temporal evolution of the “second most important” input factor. The input factors
565 (Table 2) were ranked in terms of importance based on the mean of the main effect (blue
566 straight line, Fig. 6) computed with the GP models constructed at each instant of the crisis
567 period.

568

569 *[Fig. 6 about here]*

570

571 This preliminary ranking of the input factors, only based on the mean of the main effect, was
572 assessed again in a second step taking into account the range of uncertainty associated to the
573 sensitivity measures *i.e.* using the 5% and to the 95 % quantile of the posterior probability
574 distribution associated to the main effects (black dashed line, Fig. 6). The procedure consisted
575 in qualifying the GP model as “unsure” with respect to the sensitivity measures in regions
576 where the confidence intervals of the first and second most important input factors intersect.
577 Considering the observation point 1, Fig. 6 (left) shows that for the first 150 days, coefficient
578 n_e can be identified as the “first most important” input factor with a mean of the main effect

579 constant at $\approx 20\%$, whereas the dilatancy angle Ψ can be identified as the “second most
580 important” input factor with a mean of the main effect constant at $\approx 10\%$. For the second
581 crisis period, the confidence intervals intersect and the ranking is “unsure”. Fig. 7 (left) gives
582 the mean of the main effects and the associated confidence intervals at three different steps of
583 the crisis period, namely 30 days (Fig. 7, top), 150 days (Fig. 7, middle) and 210 days (Fig. 7,
584 bottom). At 30 days, n_e can clearly be identified as the first most important input factor, but
585 the ranking of the other input factors is hardly feasible considering the intersecting confidence
586 intervals. Over time (at 150 and 210 days), the confidence intervals for all input factors
587 intersect so that the ranking is “unsure”. This result is in agreement with the low coefficient of
588 determination of the cross-validation procedure over the second half of the crisis period (Fig.
589 5, black dashed line). As a conclusion, the knowledge on the “true” simulator should be
590 increased for the second crisis time period in order to increase the predictive quality of the GP
591 model, hence to narrow the width of the confidence interval.

592

593 *[Fig. 7 about here]*

594

595 Considering the observation point 2, Fig. 6 (right) shows that before ≈ 50 days, the confidence
596 intervals intersect and the ranking is “unsure”. Over the time period after ≈ 50 days,
597 coefficient n_e can be identified as the “first most important” input factor with a mean of the
598 main effect increasing from $\approx 20\%$ to $\approx 45\%$, whereas the plastic compressibility β can be
599 identified as the “second most important” input factor with a mean of the main effect
600 approximately constant and equal to 15% . As for point 1, Fig. 7 (right) gives the mean of the
601 main effects and the associated confidence intervals for steps 30 days (Fig. 7, top), 150 days
602 (Fig. 7, middle) and 210 days (Fig. 7, bottom). It shows that over time, n_e and β can be
603 identified “with certainty” as the “first and the second most important input factors” for steps

604 150 and 210 days, but the ranking of the other input factors is hardly feasible considering the
605 intersecting confidence intervals.

606 Despite the limited number of simulator runs (30) *i.e.* the limited knowledge on the “true”
607 simulator, several conclusions can still be drawn to guide future investigations. The sensitivity
608 analysis based on the GP modelling emphasizes coefficient n_e as the “most important” *i.e.* as
609 the input factor requiring further investigations over the crisis period, whatever the part of the
610 landslide (upper or lower). In practice, the estimation of this parameter is strongly dependent
611 on the availability of lab tests at small strains, where the behaviour is truly elastic (e.g. strains
612 lower than 10^{-4}). This condition is not realized for classical triaxial tests where the accuracy is
613 not better than 10^{-3} (e.g. Biarez and Hicher, 1994) so that this parameter is usually deduced
614 using standard values estimated for analogous types of soil. Nevertheless, such an analogy-
615 based approach is hardly achievable in the La Frasse landslide case as the considered soil
616 material, being on the slip surface, is inherently heterogeneous.

617 The sensitivity analysis also outlines the plastic compressibility β as “important” for further
618 investigations in the lower part of the landslide *i.e.* where the evolution of pore pressures was
619 the most important. In practice, this parameter can be obtained from oedometer tests. No
620 further conclusions can be drawn without increasing the knowledge on the “true” simulator,
621 for the third (or lower) “most important input factor” due to the uncertainty on the GP model.
622 These conclusions are valid for the considered illustrative case especially regarding the
623 assumptions on the range of uncertainty assigned to all input factors (variation in a range of
624 25 % around the original values). Within a procedure aiming at calibrating the observed
625 displacements with the simulated ones, the uncertainty on each input factor should be
626 adequately represented making use of any kind of information related to the measurement
627 procedure of the constitutive model parameters (number of samples, estimation of

628 measurement error, possibility to construct empirical probability distribution, error of
629 calibration between observed and simulated curves, etc.).

630 **Concluding remarks and further works**

631 Landslide numerical modelling involves a large number of input factors, whose influence and
632 importance should be assessed to guide risk management and possible further investigations
633 (laboratory or *in situ*). A variance-based global sensitivity analysis (Saltelli *et al.*, 2008) using
634 the calculation of Sobol' indices (Sobol', 1993; Archer *et al.*, 1997; Sobol' and Kucherenko,
635 2005) can provide such guidelines. Nevertheless, such an analysis requires a large number of
636 direct simulations (*i.e.* simulator runs), which can be unfeasible in practice for
637 computationally intensive models (*i.e.* those characterized by computation times ranging from
638 several hours to several days). In this paper, we proposed a methodology based on Gaussian
639 Process meta-modelling to perform such an analysis using a limited number of training
640 samples.

641 The construction of the training sample is based on a space-filling approach using Latin
642 Hypercube sampling. The possible correlation between input factors is not tackled in this
643 paper and this can be further developed using, for instance, the works of Hamm and co-
644 workers (Hamm *et al.*, 2006). We presented the construction of the meta-model and how to
645 combine it with a strategy to verify the predictive quality based on a cross-validation
646 procedure. This methodology is demonstrated on a numerical model of La Frasse
647 (Switzerland) landslide (Laloui *et al.*, 2004), where the importance of the main constitutive
648 model parameters describing the slip surface material behaviour is assessed. Due to high
649 computational costs, the GP model is constructed only using 30 simulator runs, *i.e.* with a
650 limited knowledge of the "true" simulator. This induces an additional source of uncertainty
651 (referred to as code uncertainty of the meta-model) on the sensitivity measures, which is

652 tackled by treating the GP model from a Bayesian perspective: the full posterior probability
653 distribution associated with the sensitivity measures is computed and summarized by a
654 confidence interval used to outline the regions where the GP model is “unsure” with respect
655 to the sensitivity measures. When a large number of input factors (> 30) are present, the
656 Bayesian treatment of the GP model may show limitations as it is more computationally
657 demanding compared to other meta-model techniques (Storlie *et al.*, 2009). However, recent
658 works (e.g. Marrel *et al.*, 2009) pertaining to variable selection for GP model can be used to
659 overcome this difficulty. In the identified “unsure” regions, further simulator runs should be
660 carried out and the choice of the new input configurations can be guided by taking advantage
661 of the recent advances in adaptive design of experiments (e.g. Gramacy and Herbert, 2009),
662 which constitutes a possible future direction.

663 **Acknowledgements**

664 This work was funded under the BRGM’s Directorate of Research project VULNERISK. The
665 application is based on the finite-element landslide model built by the LMS of EPFL. We are
666 grateful to B. Gramacy for useful discussion on the hyper-parameters estimations of the GP
667 model. We would like to thank Dr. J. Douglas for proofreading and the two anonymous
668 reviewers for their detailed and constructive reviews.

669 **References**

- 670 Archer, G.E.B., Saltelli, A., Sobol’, I.M., 1997. Sensitivity measures, ANOVA like
671 techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation*
672 58, 99–120.
- 673 Aubry, D., Chouvet, D., Modaressi, A., Modaressi, H., 1986. Gefdyn software - Logiciel
674 d'analyse du comportement mécanique des sols par éléments finis avec prise en
675 compte du couplage sol-eau-air (Gefdyn software, Finite element analysis of soil

676 mechanical behaviour taking into account the soil-water-air coupling). Report of Ecole
677 Centrale Paris (in French).

678 Aubry, D., Hujeux, J.-C., Lassoudière, F., Meimon, Y., 1982. A double memory model with
679 multiple mechanisms for cyclic soil behaviour. In: Proceedings of the International
680 symposium on numerical models, Balkema, pp. 3–13.

681 Biarez, J., Hicher, P., 1994. Elementary mechanics of soil behaviour. Balkema, Rotterdam,
682 The Netherlands, 208 pp.

683 Bonnard, C., Noverraz, F., Lateltin, O., Raetzo, H., 1995. Large Landslides and Possibilities
684 of Sudden Reactivation. In: Proceedings 44th Geomechanics Colloquy, Salzburg.,
685 Austria, pp. 401-407.

686 Boulahya, F., Dubus, J.G., Dupros, F., Lombard, P., 2007. Footprint@work, a computing
687 framework for large scale parametric simulations: application to pesticide risk
688 assessment and management, In: Forum EGEE Enabling Grids for E-scienceE,
689 Manchester, UK, pp. 160.

690 Cressie, N.A.C., 1993. Statistics for Spatial Data, Wiley, New York, 900 pp.

691 Crosta, G.B., Clague, J.J., 2009. Dating, triggering, modelling, and hazard assessment of large
692 landslides. *Geomorphology* 103, 1-4.

693 Dupros, F., Boulahya, F., Vairon, J., Lombard, P., Capit, N., Méhaut, J.F., 2006. IGGI, a
694 computing framework for large scale parametric simulations: application to
695 uncertainty analysis with toughreact. In: Tough symposium 2006, Berkeley, 6 pp.

696 Evans, S.G., Scarascia Mugnozza, G., Strom, A., Hermanns, R.L., 2002. Landslides from
697 massive rock slope failure. In: Proceedings NATO Advanced Research Workshop on
698 Massive Rock Slope Failure, Celano, Italy, 662 pp.

699 van den Ham, G., Rohn, J., Meier, T., Czurda, K., 2009. Finite Element simulation of a slow
700 moving natural slope in the Upper-Austrian Alps using a visco-hypoplastic
701 constitutive model. *Geomorphology* 103, 136-142.

702 Hujeux, J.-C., 1985. Une loi de comportement pour le chargement cyclique des sols. In:
703 Davidovici V. (Ed.) *Génie Parasismique*, Presses ENPC, France, p. 278–302.

704 Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B. 2004. *Bayesian Data Analysis*, 2nd edn.,
705 Chapman and Hall/CRC, London, UK, 698 pp.

706 Gilks, W., Richardson, S., Spiegelhalter, D. 1996. *Markov Chain Monte Carlo in Practice*.
707 Chapman and Hall/CRC, London, UK, 512 pp.

708 Gramacy, R. B., 2007. *tgp: An R Package for Bayesian Nonstationary, Semiparametric*
709 *Nonlinear Regression and Design by Treed Gaussian Process Models*. *Journal of*
710 *Statistical Software* 19.

711 Gramacy, R.B., Herbert K.H.L., 2009. Adaptive Design and Analysis of Supercomputer
712 Experiments. *Technometrics* 51(2), 130-145.

713 Gramacy, R.B., Taddy, M. 2010. Categorical Inputs, Sensitivity Analysis, Optimization and
714 Importance Tempering with *tgp* Version 2, an R Package for Treed Gaussian Process
715 Models. *Journal of Statistical Software*, 33, 1–48.

716 Gzyl, H., 1995. The Method of Maximum Entropy. In: Bellomo, F., Brezzi, N. (Eds.), *Series*
717 *on Advances in Mathematics for Applied Sciences*. World Scientific Publishing Co, p.
718 29.

719 Hamm, N.A.S., Hall, J.W., Anderson, M.G., 2006. Variance-based sensitivity analysis of the
720 probability of hydrologically induced slope instability. *Computers and Geosciences* 32,
721 803-817.

722 Hansen, A., 1984. *Landslide Hazard Analysis*. In: Brunsden, D., Prior, D.B. (Eds.), *Slope*
723 *Instability*. Wiley and Sons: New York, pp. 523-602.

724 Hastie, T., Tibshirani, R., Friedman, J., 2002. The Elements of Statistical Learning. Springer :
725 Data Mining, Inference, and Prediction, Springer-Verlag, New York, 552pp.

726 Kennedy, M., O'Hagan, A., 2001. Bayesian calibration of computer models (with discussion).
727 Journal of the Royal Statistical Society 63 (Series B), 425-464.

728 Koehler, J.R., Owen, A.B., 1996. Computer experiment, In: Ghosh S. and Rao C.R. (Eds.),
729 Handbook of Statistics. Elsevier Science, New York, USA, 13, p. 261-308.

730 Laloui, L., Tacher, L., Moreni, M., Bonnard, C., 2004. Hydromechanical modeling of crises
731 of large landslides: application to the La Frasse Landslide, In: Proceedings of the 9th
732 International Symposium on Landslides, Rio de Janeiro, Brazil, pp. 1103-1110.

733 Langewisch, D.R., Apostolakis, G.E., 2010. A comparison of polynomial response surfaces
734 and Gaussian processes as metamodels for uncertainty analysis with long-running
735 computer codes. In: Proceedings of the 10th International Probabilistic Safety
736 Assessment & Management Conference, Seattle, Washington USA, 12 pp.

737 Le, N.D., Zidek, J.V., 2006. Statistical Analysis of Environmental Space Time Processes.
738 Springer Series in Statistics, New York, 341 pp.

739 Lopez-Caballero, F., Modaressi Farahmand-Razavi, A., Modaressi, H., 2007. Nonlinear
740 numerical method for earthquake site response analysis —elastoplastic cyclic model
741 and parameter identification strategy. Bulletin of Earthquake Engineering 5(3), 303-
742 323.

743 Lopez-Caballero, F., Modaressi Farahmand-Razavi, A., 2008. Numerical simulation of
744 liquefaction effects on seismic SSI, Soil Dynamics and Earthquake Engineering 28(2)
745 85-98.

746 Lophaven, S.N., Nielsen, H.B., Sondergaard J., 2002. DACE—A Matlab kriging toolbox,
747 version 2.0. Technical Report IMM-TR-2002-12, Informatics and Mathematical

748 Modelling, Technical University of Denmark, 28 pp.
749 http://www.immm.dtu.dk/_hbn/dace.

750 Marrel, A., Iooss, B., Van Dorpe, F., Volkova, E., 2008. An efficient methodology for
751 modeling complex computer codes with Gaussian processes. *Computational Statistics*
752 *and Data Analysis* 52, 4731-4744.

753 Marrel, A., Iooss, B., Laurent B., Roustant O., 2009. Calculations of Sobol indices for the
754 Gaussian process metamodel, *Reliability Engineering and System Safety* 94, 742-751.

755 Martin, J.D., Simpson, T.W., 2005. On the use of kriging models to approximate deterministic
756 computer models. *Journal of American Institute of Aeronautics and Astronautics* 43,
757 853–863.

758 McKay, M. D., Beckman R. J. and Conover W. J., 1979. A comparison of three methods for
759 selecting values of input variables in the analysis of output from a computer code.
760 *Technometrics* 21, 239–245.

761 Noverraz, F., Bonnard, Ch., 1990. Technical note on the visit of La Frasse Landslide. In:
762 *Proceeding 5th International Symposium on Landslides* 3, pp. 1549-1554.

763 Oakley, J.E., O’Hagan, A., 2004. Probabilistic sensitivity analysis of complex models: a
764 Bayesian approach. *Journal of Royal Statistical Society* 66 (Series B), 751–769.

765 O’Hagan, A., 2006. Bayesian analysis of computer code outputs: A tutorial. *Reliability*
766 *Engineering and System Safety* 91, 1290–1300.

767 R Development Core Team, 2009. *R: A Language and Environment for Statistical*
768 *Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2009. ISBN 3-
769 900051-07-0, <http://www.R-project.org/>

770 Rasmussen, C. E., 1996. *Evaluation of Gaussian Processes and other Methods for Non-linear*
771 *Regression*. PhD thesis, Department of Computer Science, University of Toronto, 132
772 pp.

773 Rasmussen, C.E., Williams, C.K.I, 2006. Gaussian processes for machine learning, MIT Press,
774 Cambridge, UK, 266 pp.

775 Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P., 1989. Design and analysis of computer
776 experiments. *Statistical Science* 4, 409–435.

777 Saltelli A., Sobol' IM., 1995. About the use of rank transformation in sensitivity of model
778 output. *Reliability Engineering and System Safety* 50, 225–239.

779 Saltelli A., Chan K., Scott E. M. (Eds.) 2000. Sensitivity Analysis. In: Probability and
780 Statistics Series. Chichester: John Wiley and Sons, 475 p.

781 Saltelli, A., 2002a. Making best use of model evaluations to compute sensitivity indices.
782 *Computer Physics Communications* 145, 280-297.

783 Saltelli, A., 2002b. Sensitivity Analysis for Importance assessment. *Risk analysis* 22(3), 579-
784 590.

785 Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D. Saisana, M.,
786 Tarantola, S., 2008. *Global sensitivity analysis: The Primer*. Wiley, Chichester, UK,
787 304 pp.

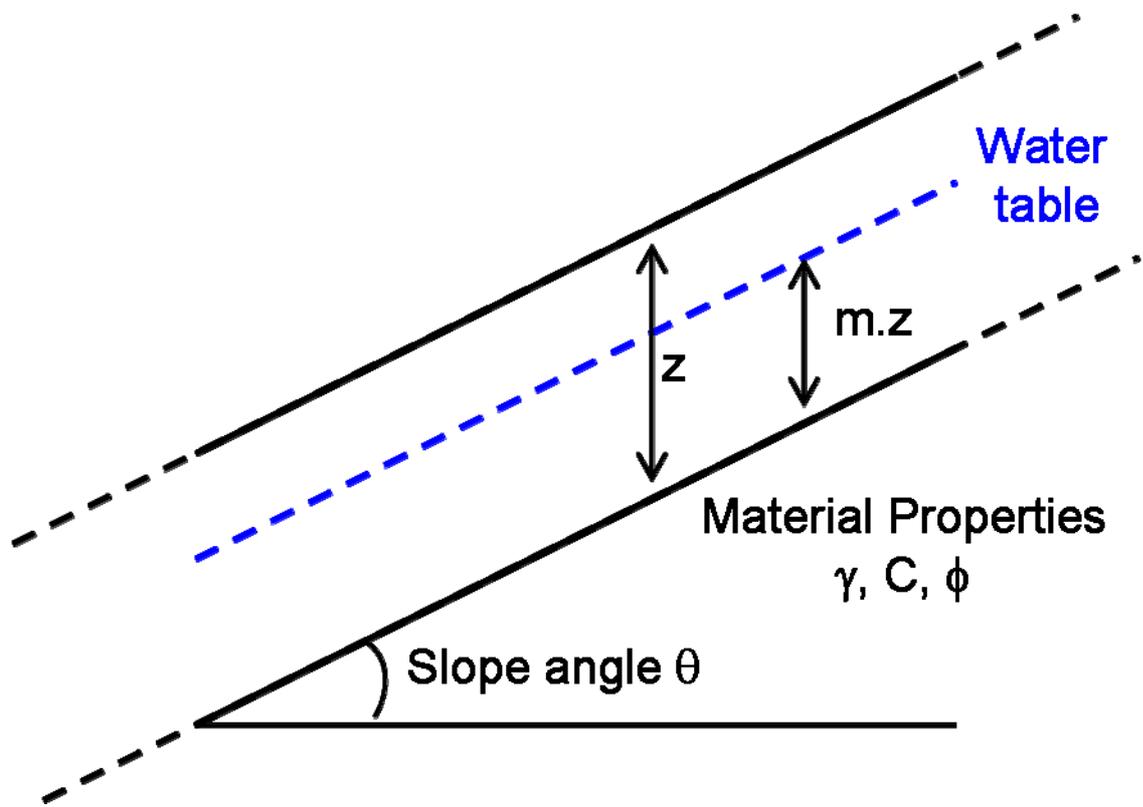
788 Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., 2010. Variance
789 based sensitivity analysis of model output. Design and estimator for the total
790 sensitivity index. *Computer Physics Communications* 181, 259–270.

791 Santner, T., Williams, B., Notz, W., 2003. *The Design and Analysis of Computer*
792 *Experiments*, Springer Verlag, New York, 308 pp.

793 Sobol', I.M., 1993. Sensitivity estimates for non linear mathematical models. *Mathematical*
794 *Modelling and Computational Experiments* 1, 407–414.

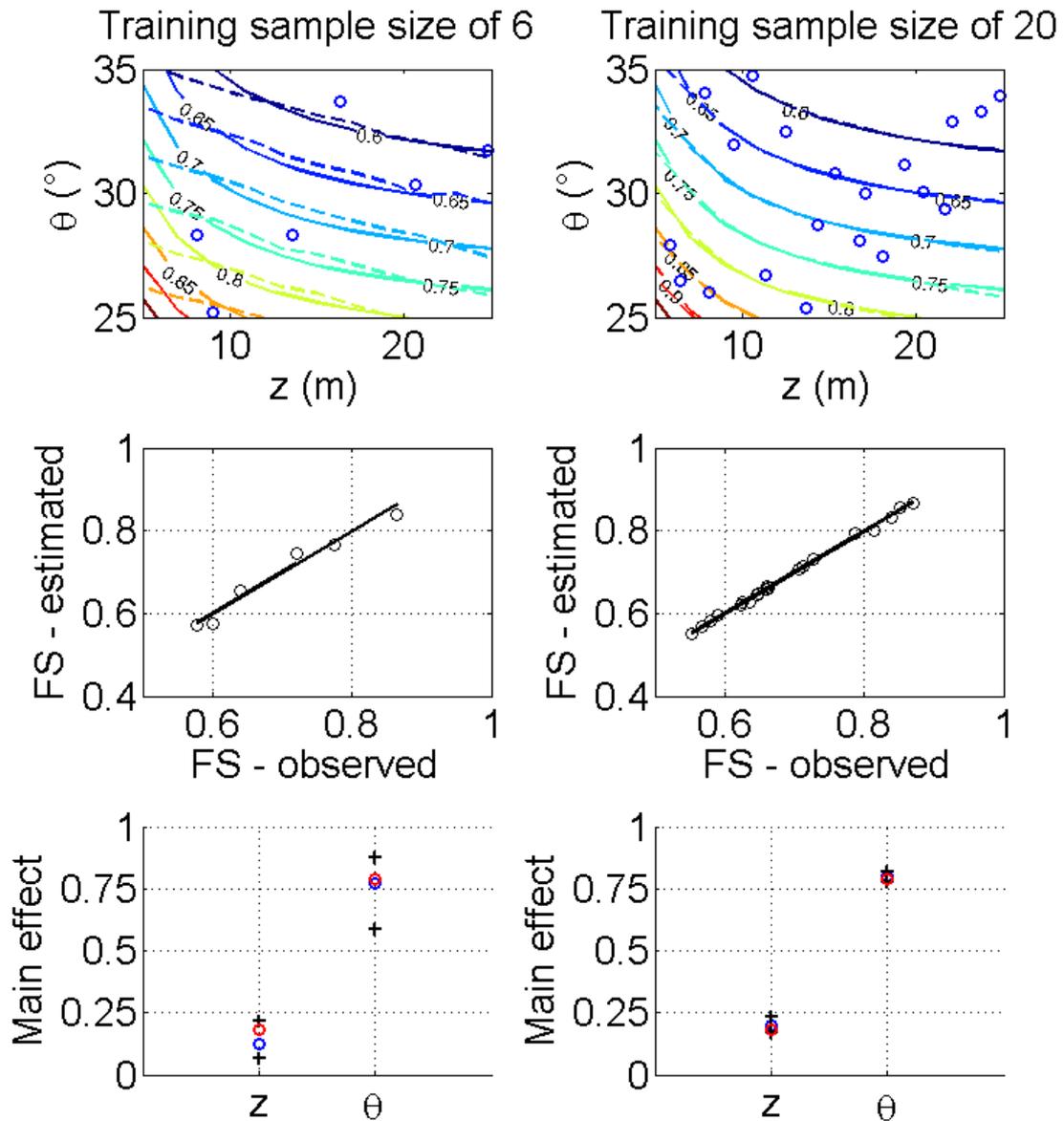
795 Sobol', I.M. Kucherenko, S.S., 2005. Global sensitivity indices for nonlinear mathematical
796 models, *Review of Wilmott Magazine* 1, 56–61.

- 797 Stein, M.L., 1999. Interpolation of Spatial Data. Springer: Some Theory for Kriging, New
798 York, 247 pp.
- 799 Storlie, C.B., Swiler, L.P., Helton, J.C., Sallaberry, C. J., 2009. Implementation and
800 evaluation of nonparametric regression procedures for sensitivity analysis of
801 computationally demanding models. Reliability Engineering and System Safety 94,
802 1735–1763.
- 803 Von Mises, R., 1964. Mathematical Theory of Probability and Statistics. Mathematical
804 Theory of Probability and Statistics, Academic Press, New York, 694 pp.



806

807 Fig. 1: Schematic representation of the infinite slope model (adapted from Hansen, 1984).



808

809 Fig. 2: Top: comparison between the true values and the estimates of the factors of safety FS

810 using the GP model constructed with a training sample size of 6 (left) and of 20 (right). The

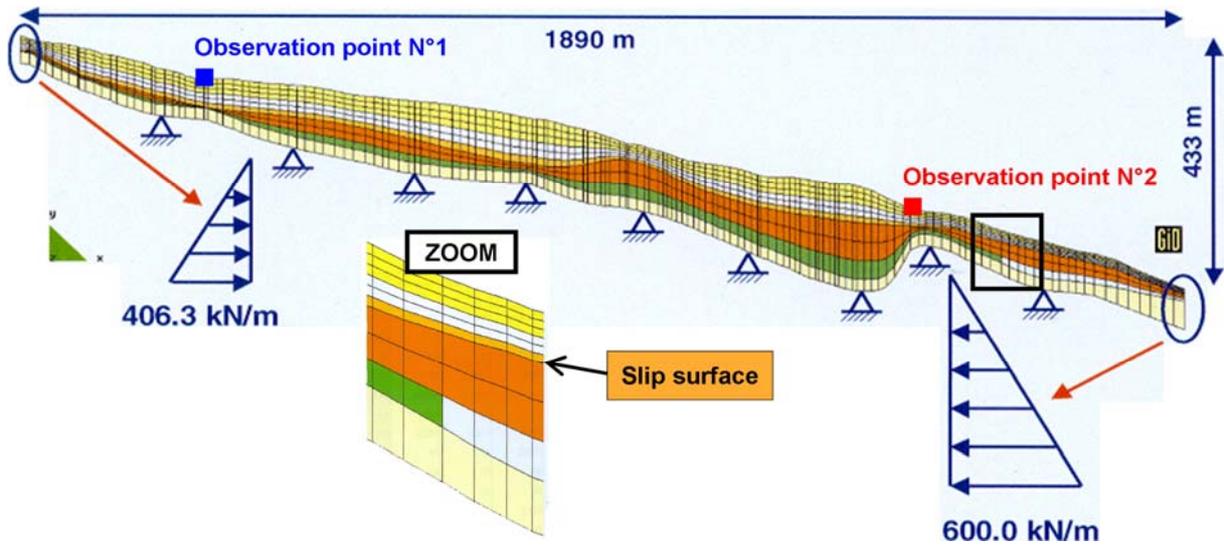
811 training input configurations are represented by blue dots. Middle: comparison between the

812 observed FS and the estimates within a “leave-one-out” cross validation procedure. Bottom:

813 comparison between the true values (red dots) and the estimates of the main effects for the

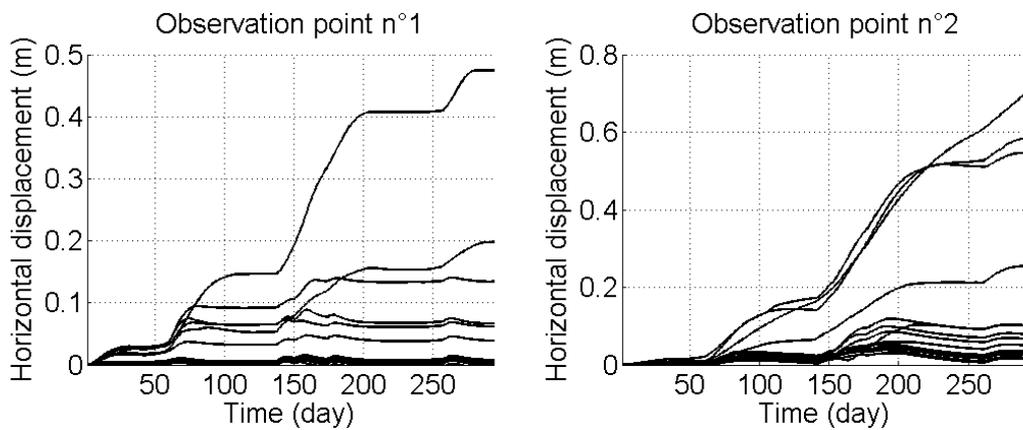
814 slope thickness z and of the slope angle θ (blue dots). The bounds of the confidence intervals

815 associated with both GP models are represented by black cross-type markers.



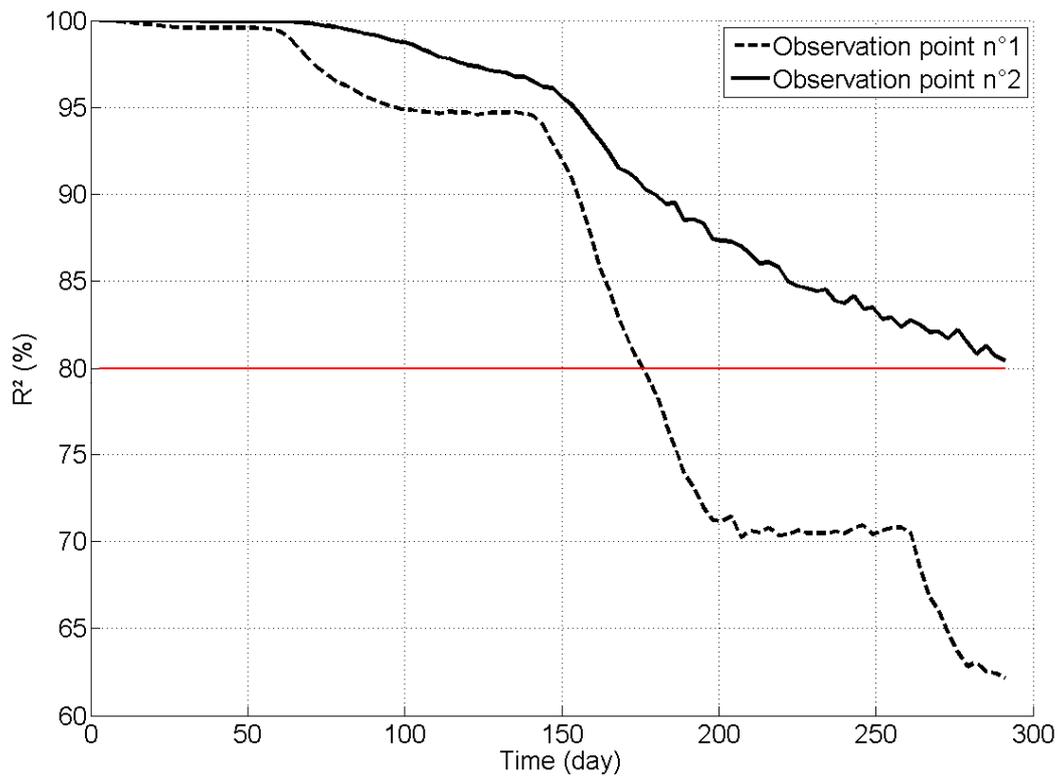
816

817 Fig. 3: Overview of the landslide numerical model (adapted from Laloui *et al.*, 2004). The
 818 slip surface is outlined by the light coloured orange surface. The observation point 1 (in the
 819 upper part of the landslide) and observation point 2 (in the lower part of the landslide) used
 820 for the sensitivity analysis of the horizontal displacements are respectively outlined by a blue
 821 and a red coloured square-type marker.



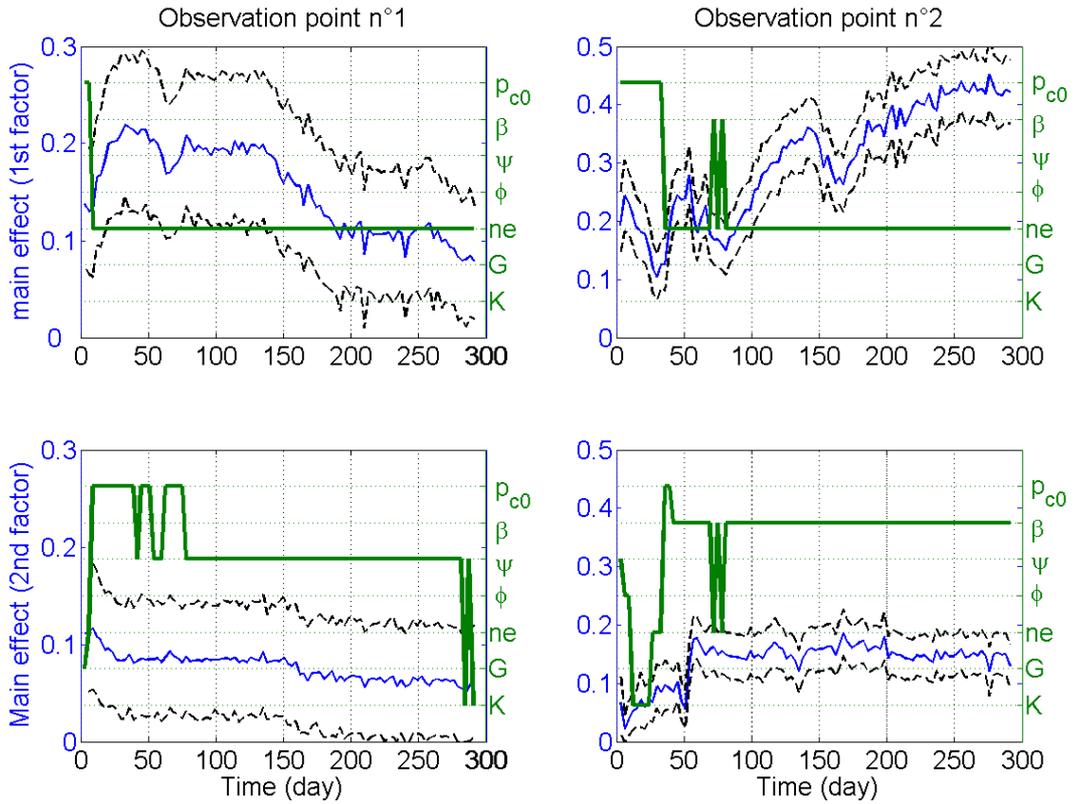
822

823 Fig. 4: Temporal evolution of the training samples corresponding to the horizontal
 824 displacements (m) calculated for 30 different input configurations of the Hujoux law
 825 parameters (at the observation point 1 in the upper part of the landslide (left) and at the
 826 observation point 2 in the lower part of the landslide (right)).

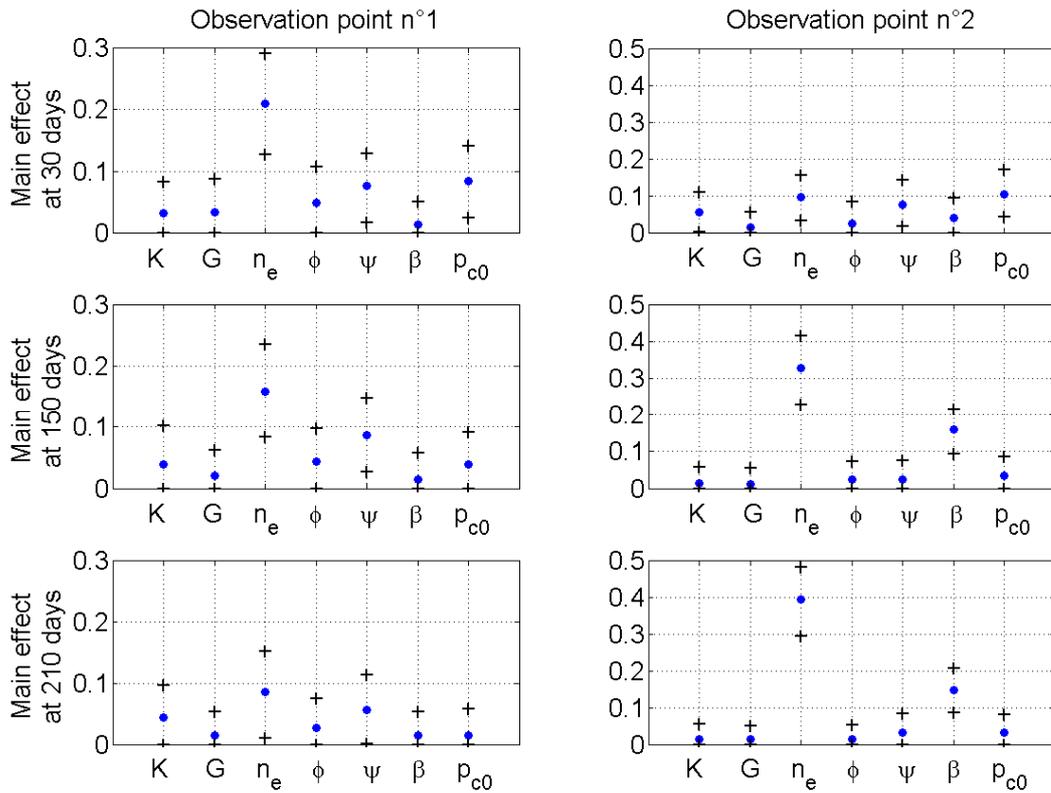


827

828 Fig. 5: Temporal evolution of the coefficient of determination R^2 for the “leave-out-out” cross
 829 validation procedure of the GP models constructed at each instant of the crisis period at the
 830 observation point 1 in the upper part of the landslide (black dashed line) and at the
 831 observation point 2 in the lower part of the landslide (black straight line). The threshold of
 832 80 % indicating a “satisfactory” predictive quality is outlined by a horizontal red straight line.



833



834

835 Fig. 6: Temporal evolution during the crisis period of the mean of the main effects (blue
 836 straight line) at the observation point 1 in the upper part of the landslide (left) and at the

837 observation point 2 in the lower part of the landslide (right) for the first (green straight line,
838 top) and the second (green straight line, bottom) “most important” input factor. The black
839 dashed lines represent the 5% and the 95 % quantile.

840

841 Fig.7: Mean of the main effect (blue dots) for each input factor of the slip surface constitutive
842 law at different instants of the crisis period (30 days (top), 150 days (middle) and 210 days
843 (bottom)) at the observation point 1 in the upper part of the landslide (left) and at the
844 observation point 2 in the lower part of the landslide (right). The bounds of the confidence
845 intervals (5% and 95 % quantile) are outlined by black cross-type markers.

846

847 **TABLES**

848

849 Table 1: Comparison between the “true” and the estimates of the main effects for the infinite
 850 slope analytical model. μ corresponds to the mean of the main effect computed with the GP
 851 model. CI corresponds to the confidence interval defined by the 5 % and the 95 % quantile
 852 computed with the GP model.

Input factor	True model	GP model constructed with 6 training samples	GP model constructed with 20 training samples
Thickness z (m)	18.41 %	$\mu=12.74$ % $CI=[6.48 ; 21.73]$ %	$\mu=20.02$ % $CI=[16.90 ; 23.29]$ %
Slope angle θ (°)	78.76 %	$\mu=77.01$ % $CI=[58.41 ; 87.43]$ %	$\mu=79.77$ % $CI=[77.90 ; 81.68]$ %

853

854 Table 2: Range of values for the slip surface properties of the La Frasse landslide (variation in
 855 a range of 25 % around the original values given in Laloui *et al.*, 2004)

Input factor	Vol. comp. mod.	Shear mod.	Non-linearity coeff.	Internal friction angle	Dilatancy angle	Plastic comp.	Initial critical pressure
Symbol	K	G	n_e	ϕ	Ψ	β	p_{c0}
Unit	MPa	MPa	-	°	°	-	MPa
Lower value	180	83.25	0.225	19.125	14.25	20.625	0.375
Upper value	300	138.75	0.375	31.875	23.75	34.375	0.625

856